Pacific Northwest
NATIONAL LABORATORY

PNNL-32684

DeepCare: Improving Patient Care using Deep Learning on Electronic Health Records

February 2022

Khushbu Agarwal
Sutanay Choudhury

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

PNNL-32684

DeepCare: Improving Patient Care using Deep Learning on Electronic Health Records

February 2022

Khushbu Agarwal
Sutanay Choudhury

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Project Description

Develop novel disease prediction models using deep learning techniques on Electronic Health records (EHR) data, potentially enabling personalized, precision medicine for patient populations.

Abstract

Coordinating patient care using electronic health records (EHR) data presents an exciting but formidable opportunity in data extraction, analysis and modeling. Traditional methods use a manual feature driven approach to model patients with age, family history and symptoms to predict disease outcomes. We propose a novel approach to model patients based on their streaming electronic health records data combined with information from medical knowledge bases, which has been gained over years of medical research. Using a combination of representation learning and long short-term memory (LSTM) networks we plan to model patient evolution over time, leading to more accurate and individualized predictive models for patient's diseases. Our approach will be transformative in providing critical decision support for patient care, enabling accurate understanding and evolution of diseases in patients.

Results and Accomplishments

We developed methods for embedding learning of medical concepts by using graph-based representation of SNOMED-CT, a widely popular knowledge graph in the healthcare domain with numerous operational and research applications. Our work included 1) Identification and extraction of medical concepts relevant to predictive learning on Electronic Healthcare Records(EHR), 2) Exploration of state of the art embedding learning methods, including Euclidean space methods such as (Node2vec and Metapath2vec) and hyperbolic space methods (Poincare) for embedding learning of the medical concepts using their graphical properties 3) Development of deep learning methods for integrating the medical concept representation into temporal patient state prediction models 4) An empirical analysis of various embedding methods, including the evaluation of their performance on multiple tasks of biomedical relevance (node classification, link prediction, and patient state prediction).

Our results show that concept embeddings derived from the SNOMED-CT knowledge graph significantly outperform state-of-the-art embeddings, showing 5-6x improvement in "concept similarity" and 6-20% improvement in patient diagnosis. The work was published in a KDD workshop (https://arxiv.org/pdf/1907.08650.pdf) and selected for spotlight presentation for Health day (https://dshealthkdd.github.io/dshealth-2019/). The code and the generated medical concept graph with embeddings was released as open source for use by the medical informatics community.(https://gitlab.com/agarwal.khushbu/Snomed2Vec/-/tree/master)

The project also led to sponsor funding from Veteran Affairs for development of predictive models for patient health (1.2M over 2 year to date) and setting up active collaborations with Stanford University, Center of population health sciences.

Acknowledgements

This research was supported by the Physical and Computational Sciences Directorate Mission Seed, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov